

# MULTILINGUALISM IN INFORMATION RETRIEVAL SYSTEMS THE NEXT CHALLENGE<sup>1</sup>

**Pat Riva**

*Associate University Librarian,  
Collection Services  
Concordia University, Montreal,  
Quebec, Canada*

**KEYWORDS:**

*multilingual catalogues, bilingual catalogues, language of cataloguing, multilingual name authority files*

**ABSTRACT**

*A multilingual information retrieval system is one that serves, from a single system, a user population which as a whole is functioning in multiple languages. The multilingual design of the information retrieval system compensates for the differences in language preference of the end-users. The system would ideally be as good in each of the languages it offers. Many layers in a catalogue have linguistic orientation and need to be considered when reorienting a single-language catalogue into a multilingual catalogue. These include: the user interface, data display conventions, and all metadata, from statements transcribed from a resource to cataloguer-supplied data, access points, and subject languages. Despite known issues, interfaces that can be switched into multiple languages are widely implemented. Metadata is not always language-neutral. A choice of language is implicit in aspects of our existing cataloguing code or guidelines.*

1 Expansion of a presentation given at the colloquium *Diverse Perspectives on Bilingualism*, October 12, 2018, Concordia University, Montreal, Canada.

*Decision criteria that rely on determining well-established forms of names or titles in a language preferred by the cataloguing agency result in descriptions that are oriented to users of a particular language. Preferred forms are controlled by name authority files, which can be leveraged to adapt the bibliographic data displayed to the end-user. Approaches for subject access are direct linking of the different subject schemes, or using a classification as a common linking hub to navigate between languages. Despite the advances that have been made, there are still basic challenges to providing fully functional bilingual and multilingual systems. Partial multilingualism may still be more useful than not attempting it at all.*

## The global context of information retrieval

Information retrieval is a core interest in library science. It could be described as the process of obtaining bibliographic resources relevant to an information need from an organized collection of information resources. Many methods for organizing collections and providing metadata to facilitate retrieval tailored to specific contexts are in use. Information retrieval systems have usually been focused on being usable in a single language. The expectation behind this is that the intended users of a system will be homogeneous and share a single preferred language for accessing the system. When a system is intended to serve the users of a single library, this would in most cases be a reasonable assumption. If a library serves a single institution or a single community, these users could reasonably be expected to have some characteristics in common, such as a common language.

This assumption may come to be increasingly questioned. Communities or user groups are not necessarily linguistically homogeneous, for example in multilingual countries or border regions. Additionally, cloud-based systems make it increasingly feasible to expand the scale of the user group served by a single system. Information retrieval systems now frequently serve cooperating groups of institutions or consortia over larger geographic regions. The geographic distance between institutions is becoming less of a barrier in forming these consortia, considering that many of the resources offered by the participating libraries may be digital, and so can be accessed by users wherever they may be (as long as appropriate licenses are in place). Even physical resources can, in many cases, be shared rapidly among groups of cooperating libraries, making it interesting and relevant for users to discover resources outside their direct geographic area.

### What do we mean by a bilingual or multilingual user population?

Describing a user population as bilingual or multilingual refers to a characteristic at the level of the whole user population, not necessarily a characteristic of all the individuals within the population. A bilingual user population may include individuals who are effectively unilingual in each of the languages along with individuals with varying levels of ability in both of the languages. Linguistic research indicates that although bilingualism is quite common world-wide, the proportion of bilingual people who are “balanced” bilinguals (those equally

fluent in both of their languages) is the minority (Grosjean, 1982, pp. 230-240). More common are bilinguals who are functionally bilingual, meaning that they regularly use two languages. Fluency is a continuum. Domains of use for each of a person's languages also affect a person's language fluency. A bilingual person may use each language in specific contexts, such as when speaking with certain people, or for certain activities. One of the languages may be the person's dominant language, the one they feel most comfortable with. This language may be the same language as their preferred language for information searching, but in some cases it may not, as the individual may favour a different language for their professional work than for their family and social life.

### **Defining a multilingual information retrieval system**

A multilingual information retrieval system is one that serves, from a single system, a user population which as a whole is functioning in multiple languages. In a case where all members of the user population are fully functional in a single language (the same language), then the information retrieval system could be unilingual and still serve them all adequately. Bilingual or multilingual systems are only needed when this is not the case. The multilingual design of the information retrieval system compensates for the differences in language preference of the end-users.

A multilingual design could be a system that is used in one language at a time, but is dynamically switchable to a different language. Alternately, one could imagine a system functioning simultaneously in several languages at once. However, displaying multiple languages at once will rapidly run into issues of scalability, if only for screen space and clarity of presentation. Considering that the user population is likely to have varying comfort levels with the languages enabled in the system, transforming the interaction into another language is overall going to provide a better experience than always presenting all the languages.

The ultimate goal of a bilingual or multilingual information retrieval system of this kind is to provide a fully equivalent experience to the end-user whichever of the available languages that user chooses to use to interact with the system. The system would ideally be as good in each of the languages it offers. For an information retrieval system to be considered good it should be retrieving relevant resources with strong scores in both precision and recall.

## Making the catalogue multilingual

For our purposes we will define a multilingual catalogue as a single database and system that can present an end-user experience appropriate for retrieval of bibliographic resources in multiple single languages.

Making the catalogue usable in multiple languages will require a combination of making some aspects switchable into multiple languages, making some aspects multilingual (functional in several languages all at once), and making some aspects language-neutral so that they can be plugged as is into shells of different languages. Pragmatically, it will also require making the most of the available metadata, even that created using unilingual conventions. Designers will need to look to various strategies to compensate for the data, including creative search features, linkages in authority files and expansions of controlled vocabularies or coded data from multilingual lists of captions. Good candidates for this last strategy might include relationship designators, language codes, or geographic area codes.

Many layers in a catalogue have linguistic orientation and need to be considered when reorienting a single-language catalogue into a multilingual catalogue. A by-no-means complete list includes:

- the user interface (including all commands, menus, messages);
- data display conventions (such as layout, labels, or displayed equivalents for coded data);
- the data itself (both transcribed data from a resource and cataloguer-supplied data such as free-text notes);
- access points (including those for agents, works and expressions), and
- subject languages (subject headings or thesauri).

A catalogue is populated with metadata describing and providing access points for discovery of bibliographic resources. Bibliographic resources intersect with language: the bibliographic resources are written (or spoken) in one or more languages and describe themselves in one or more languages. However, description conventions are intended for users of a particular language. A choice of language is implicit in aspects of our existing cataloguing code or guidelines. How can these different aspects be approached in making a multilingual catalogue?

## Interface languages

The first step to making a catalogue usable in multiple languages is to translate the user interface. This means having full language versions of all text seen by end-users, including the help system, and any messages or responses given to end-users. Although a well-understood need, this isn't always easy to achieve. For instance, the text on a button may not fit when rendered in the target language, if that concept is expressed by a longer term in the target language than in the source language. Thinking ahead in the design phase to making an interface translatable can make it easier to make the interface look good in a target language.

Full localization includes more than simple translation of text. Languages differ in such things as the alphabetical order of the script, how diacritics should be sorted, interpreting symbols or layout conventions. Many small details trigger unconscious linguistically-based expectations in the end-user. If these conventions are not respected, the interface will not feel natural in the translated language.

A whole other level of complexity is added by different scripts, and particularly the directionality of scripts. A great deal has been written on these topics, which are fundamental to building any comprehensive multilingual system (Aliprand, 2005). Despite such issues, adding more options for interface languages is relatively easy compared to other aspects and is now commonly done. But this isn't enough to fully render a system usable in another language.

Data labels may be interpreted by the end-user as being part of the data, or as being part of the interface, depending on their position and layout. When labels are set off from the data that follows them, such as in a labelled bibliographic description, they are clearly separate from the data and make sense in the interface language. Coded data expanded for display within data blocks can appear to be part of the data and needs to follow the metadata language.

## The resources in the collection

A collection intended for a bilingual or multilingual user population will be expected to contain bibliographic resources in each of these languages, and often more. Some resources are themselves multilingual. Since each end-user varies

in their ability to make use of resources in different languages, the language or languages of the content of the resources needs to be recorded accurately and presented to the user as a means of selecting relevant resources.

The languages in which a resource presents itself, usually, but not always, follow the languages of the content. The principle of representation indicates that this information is to be preserved and recorded in the language in which it is found. In a multilingual catalogue, parallel data gains in importance. Recording data in all the languages it is presented in is one way to ensure that descriptive metadata is available in multiple languages for the end-user to see and so recognize that a given resource does have content in a language usable by the user. However, parallel language data has often been abridged in resource descriptions intended for a unilingual catalogue, which may limit the discoverability of language-appropriate resources.

These transcriptions remain fixed in the language found on the resource, regardless of the language of the user interface or the language of the cataloguing agency. As a result, record display naturally has a multilingual flavour. Titles, parallel titles, series and other transcribed data appear in the languages found on the resource, none of which might match the language of the user interface, nor the language or languages of the subject vocabulary.

### **Managing languages separately**

The Library and Archives Canada *Bilingual Cataloguing Policy* (2003) presents one approach for a national bibliographic agency in a bilingual country. A very high-level summary of the policy is that French-language bibliographic resources are catalogued in French, while English-language ones are catalogued in English, as are any resources in any other language. Bilingual resources are catalogued twice, once in each language. Both French-language and English-language subject headings are provided in all records. The records produced under this policy can be derived and easily integrated into either English-language or French-language catalogues, serving unilingual client libraries well. But within the combined catalogue, this results in language silos for the end-user and is oriented to the idea that each user is unilingual and will not use resources in the other language.

However, a key characteristic of multilingual users is that they are able, and for many purposes, willing, to use resources that are in any of the languages

they know. In language learning it is recognized that it is easier to understand a language than to produce in it. The effect on searching is that devising a correctly spelled search query in one's non-dominant language may be difficult, but recognizing a relevant resource although the title is in that language is much easier. And the resource may still meet the need, as reading ability is generally stronger than production for most bilingual and multilingual people. The interest for resources written in multiple languages is not only from researchers in academic settings. International students, immigrants, travellers, second-language learners, and many other people in a multilingual community using many different libraries, will have similar needs. Cataloguing policies that create language silos put the burden of searching on the bilingual end-user who must repeat the search in each language to retrieve all the relevant resources that are held.

## Description conventions

For use in a multilingual information retrieval system, it is not bilingual or multilingual resource descriptions that are needed, rather language-neutral descriptions that could be used or adapted in a simple way to be appropriate in conjunction with the interface language. By “being appropriate with” we mean that in some cases the data is transformed to match the interface language, while in other parts of the description, sticking with the language of the resource itself. This is really language-neutral description, not a single description that is simultaneously in more than one language.

The biggest hurdles in making resource descriptions language-neutral are found in the cataloguing conventions themselves. The aspects of resource descriptions that are dependent on the choice of a preferred language by the agency are not obvious at first glance; simple descriptions of straightforward resources do appear to be language-neutral. And the parts of the description that are not language-neutral mainly seem as though they could be transformed by machine by maintaining simple equivalence tables for those lists of controlled terms (terms specified in the cataloguing instructions such as “pages”, “volumes”, “illustrations”, “maps”, or the controlled terms for content type, media type and carrier type). Cataloguer supplied free-text notes are another matter. Only some notes follow simple predictable patterns that could be mapped to another language.

Much of the description consists in recording manifestation statements taken from the resource itself. The principle of representation, found in the IFLA Statement of International Cataloguing Principles, naturally leads to following the language or languages those statements are presented in. This remains constant, whatever the language of the user interface. The choice of which statements to transcribe does, however, depend on the determination of a preferred source of information. In RDA, *Resource Description and Access*, a process for deciding on the preferred source of information in those cases where there is more than one candidate, is laid out in instruction RDA 2.2.3 (More than one preferred source of information). The basic instruction is to “use the first occurring of these sources”, with a few special cases. One special case is when the sources are in different languages or scripts. RDA 2.2.3.1 (Preferred sources of information in different languages or scripts) lays out an order of preference for choosing a source in various linguistic situations. This seems to be an objective process; the whole decision-tree is dependent on characteristics inherent in the resource itself, and should lead to the same choice of preferred source of information for all cataloguing agencies. Until we get to the very last situation, RDA 2.2.3.1 f):

the source in the language or script preferred by the agency preparing the description, if the resource is formatted tête-bêche, as a head-to-head bound monograph, or as a head-to-tail bound monograph.

In this case, a characteristic of the agency is brought to bear in the interpretation of the physical resource, as there is no inherent preference, in the presentation of a tête-bêche bilingual monograph, for one or the other of the languages. But this changes everything about the manifestation statements that are transcribed. And the impact of this decision goes beyond the description of a specific manifestation, as by specifying the title proper that is recorded, this choice may even determine the preferred title that is assigned to the work, preventing even the work-level identification from being language-neutral.

Generally, the principle of representation will result in language-neutral choices for the preferred forms of names of agents in responsibility relationships. In most cases, the form chosen would follow the usage of the entity itself, determined by various criteria. However, when the entity uses multiple forms of name in different languages, then the language of the cataloguing agency is a deciding factor under some conditions.

For example, in RDA 9.2.2.5.2 (Language), for a person whose name consists “of a given name and/or a word or phrase associated with the person”, the preferred form will be “a well-established form of the name in a language preferred by the agency creating the data”.

Or, for corporate bodies with more than one official language, RDA 11.2.2.5.2 (Language) states:

“If there is more than one official language and one of these is a language preferred by the agency creating the data, choose that form as a preferred name.”

And similarly, at RDA 11.2.2.5.3 (International Bodies):

If the name of an international body appears in manifestations associated with it in a language preferred by the agency creating the data, choose that form as a preferred name.

These guidelines result, in an English language catalogue, in these choices:

- Francis of Assisi, instead of Francesco d’Assisi;
- Canadian Committee on Cataloguing, instead of Comité canadien de catalogage;
- International Federation of Library Associations and Institutions, instead of Fédération internationale des associations de bibliothécaires et des bibliothèques.

The little phrase “a well-established title in a language preferred by the agency” appears also in some special cases for the choice of a preferred title of a work, such as at the exception for Classical and Byzantine Greek Works at RDA 6.2.2.5 (Works Created before 1501).

All these factors make it complex to imagine how a single bibliographic description package could be designed that is in all cases both coherent and truly multilingual or entirely language-neutral. The intention in this section is not to provide a complete listing of dependencies on agency language in any specific set of descriptive cataloguing guidelines, but merely to identify some patterns that will lead to inserting the language of the cataloguing agency into the resulting descriptive metadata.

## Name authority files and language

Many of the cases that prevent a traditional bibliographic description from being language-neutral involve the choice of preferred form of name for agents or works and expressions that is displayed in the description. As these forms are controlled by name authority files, leveraging the authority files to adapt the bibliographic data displayed to the end-user according to the desired linguistic context provides an interesting opportunity. Various methods for this have been tried.

One approach that is built into the MARC 21 format is the use of 7XX Heading Linking Entry fields in the authority format. One of the use cases for these fields specifically mentioned in the MARC 21 documentation is to link:

Equivalent names in a multilingual thesaurus. Example: Library and Archives Canada English heading *Francis, of Assisi, Saint, 1182-1226* and Library and Archives Canada French heading *François, d'Assise, saint, 1182-1226*.

In the example given from the Library and Archives Canada, Canadiana<sup>2</sup> name authority file shows how the preferred form of name for the same person differs depending on whether the language of cataloguing is English or French. In this type of structure, two authority records are created, one for each language of cataloguing. These records are then linked using reciprocal 700 fields pointing to the other record. The link can be implemented to switch a display form in a bibliographic record display and also to enrich name indexes by using the information from both authority records, whichever of them is actually directly linked to the form found in the bibliographic file. This method allows the two language files to function independently or to be used together. However, scale is an issue. If reciprocal links need to be established between three, four or even more language-based name authority files, the increasing maintenance burden becomes prohibitive.

Another approach is to host multiple preferred forms within a single authority record. This was pioneered in Switzerland to handle the four national

2 In 2018 the Canadiana name authority file was split by language. The English portion of the file was added to the LC/NACO name authority file, while the French portion of the file became an independent file under the name Canadiana. However, the 7XX fields remained in the records to provide the linking function. Source: <https://www.bac-lac.gc.ca/eng/services/national-union-catalogue/Pages/questions-answers-contract-oclc.aspx#what-options> (What will happen with Canadiana Authorities in English and in French?)

languages from a single national authority file (Lehtinen and Clavel-Merrin, 1995). This approach has been extended in a multiscrypt multilingual context at the National Library of Israel (Cohen, 2017), and by the libraries in Nunavut, Canada (Rigby, 2008), among others. Special purpose, locally-maintained additions to standard MARC 21 are used in the Swiss and Israeli files to label the access points with the script or language for which each is intended. In this approach, the rest of the record is shared by all the languages or scripts, meaning that variant forms only need to be recorded one time. The preferred forms in other languages can be indexed as additional variant forms. In this approach it is important to understand whether the goal is to label the access point with the language it is in, or to label it according to which of the supported catalogue languages it is considered appropriate for. This distinction comes into play when the content of the access point is in an additional language, one that is not supported by the file. This complexity has resulted in a reluctance to enable this approach in a standard MARC 21 authority file. Several MARC discussion papers, the latest being 2001-DP05, have been discussed without resulting in any proposals. While this approach has been adopted in contexts with four languages or scripts, extending such a file indefinitely, to a large number of languages, would become increasingly difficult. One practical difficulty would be finding cataloguers able to work simultaneously in so many languages.

A third approach involves mapping distinct authority files virtually, without creating hard-coded links stored in each of the separate authority records. This is the approach taken in the Virtual International Authority File (VIAF), which is the most extensive example of a multilingual and multiscrypt file. There are over 50 participating files as of this writing, in many languages and scripts, as well as using different cataloguing rules and conventions. VIAF has a good success rate in matching authority records representing the same real world entity, regardless of the differences between preferred forms found in the authority records. This is accomplished by exploiting all sources of information found in the authority records and in the linked bibliographic records provided by participants. VIAF shows the strong potential of this approach, which so far has not been integrated into library systems to allow for dynamic display of appropriate forms to end-users.

## Subject access

Probably the most progress in providing multilingual access has been made in the sphere of subject access. There are two main approaches. Direct linking of the different subject schemes or using a classification as a common linking hub to navigate between languages.

For textual subject languages, one approach is cross-linking projects between subject files. The Swiss-led project MACS (Multilingual Access to Subjects) (Clavel-Merrin, 2004; Landry, 2000; Clavel-Merrin, 1999) is the classic project that comes to mind, with cataloguer-validated linking of English, French and German subject headings from the LCSH, Rameau and SWD (now part of the GND) files respectively. The National Library of Israel also hosts a bilingual LCSH and Hebrew language subject file (Cohen, 2017). This cross-linking works best when the subject languages have similar scope and structure, but has also been accomplished with schemes of differing structures (Mayr and Petras, 2008). The search interface can make use of these links to include in the results set bibliographic records that are indexed using any of the terms marked as other-language equivalents in any of the relevant vocabularies, as well as by exploiting the within-language cross-references and variant terms. As with linking of name authority files, a lot of intellectual labour is needed to correctly match the terms. In some cases, there may be no match possible, or only broader term matches may exist, and so precision of retrieval can be reduced.

Recording these links in the subject authority records will have all the same issues of scalability as with name authority files. To reduce the links needed, one can link each language only to a central hub or switching language. A relatively neutral choice for this purpose can be to use a classification scheme that has already been translated into multiple languages. For instance, the current edition of the Dewey Decimal Classification has been fully translated from the original English, into French, German, Italian, Norwegian, and mostly into Swedish. This could be implemented to facilitate multilingual lookup of the captions and scope notes for any notation assigned in the bibliographic record. Here, once the user identifies a classification of interest (this could be by other search keys, or subject headings, or via the entry terminology in the captions and scope notes in the user's preferred language), the retrieval system can retrieve bibliographic records indexed with that classification, regardless of the language of cataloguing of those bibliographic records. How well this works does depend on how amenable the classification scheme used as a hub

is to being adapted to the target languages and cultures, which is in itself not a negligible issue.

## Multilingualism in cataloguing tools

Similar approaches for creating multilingual systems can be seen in some of the cataloguing tools that are intended to serve cataloguers working in different languages using a common tool. These systems can provide interesting parallels and possibly transferable solutions to the goal of designing a multilingual catalogue.

The WebDewey interface by Pansoft is very successful in its bilingual English-French Canadian implementation, available since 2015 (Werling, 2015). Both the interface and the data behind the classification have been fully mapped between languages, making the cataloguer's user experience in each language truly equivalent. The language of interface is fully independent of the language of the classification data, and the cataloguer is allowed to select these languages independently. The languages can be switched at any point in the use of the system; language selection is not just a one-time choice that must be made when starting a session. Maintenance of the linguistic mappings in a dynamic classification that is continually developing remains a challenge. Having the system gracefully handle the situation of a switch in the data language at a point where the linguistic equivalent is not available (because it is not yet translated), required some thought, as the classification number may not yet be transferred to the translation language.

The original ("classic") RDA Toolkit also followed a design separating the language of the interface from the language of the RDA text. The cataloguer could switch the language of the RDA text dynamically from any section, and also display two languages in a split-screen mode. And this independently of the choice made in the parameters for the language of the interface.

In these tools, the multilingualism has an important function in supporting the translation teams that keep these systems up-to-date. However, cataloguers working in bilingual libraries can also find it useful to switch languages in their work. For instance, when applying the Dewey classification, it can be helpful to verify the scope of a class number by reading the scope note in the language of the resource being classified.

## Concluding thoughts

Language is inextricably bound up in communication, and so is fundamental to information retrieval. As we move away from language silos and expand our information horizons, the ability to provide end-users with an appropriate experience in more than one language becomes a higher priority.

Despite the advances that have been made in localizing interfaces and linking subject languages and authority files, there are still basic challenges to providing fully functional bilingual and multilingual systems. Some of these challenges stem from system design and can be addressed by consciously designing for bilingualism, rather than attempting to retrofit unilingual systems. But is that enough? Can bilingualism readily scale up to multilingualism? If a retrieval system has been successfully made bilingual, is adding a third or fourth language, or more, just a simple extension, a matter of having the appropriate language content? I suspect this is not the case, and that designs will need to be explicitly multilingual from the outset.

Partial multilingualism may still be more useful than not attempting it at all. In this situation, it is important to consider how systems should behave when some of the data is not available in one or more of the supported languages. Designing the system to fail gracefully, so that the results are still as useful as possible to the end-user.

This overview of selected aspects and issues in making catalogues multilingual makes no claim to be a complete review of previous work in this area, just to highlight some important areas that will need to be addressed.

A system must have data, else it is an empty shell. In my view, the greatest challenge lies in adapting our descriptive standards to create language-neutral, or at least language-adaptable, resource descriptions. Language of cataloguing is inextricably bound to some aspects of current description conventions. This may be unavoidable to some extent, and so we will need to devise other strategies for making a single description adaptable to multiple language contexts.

In conclusion, adapting catalogues for multilingual information retrieval is going to require some rethinking at a fundamental level. No single technique will accomplish the full goal. This is proving to be the next big challenge and essential for heritage institutions functioning at an increasingly global level.

## REFERENCES

- ALIPRAND, J.M. (2005). Scripts, languages and authority control. *Library Resources & Technical Services*, 49, 243-249.
- CLAVEL-MERRIN, G. (1999). The need for co-operation in creating and maintaining multilingual subject authority files. In *65th IFLA Council and General Conference, Bangkok, Thailand*, <http://archive.ifla.org/IV/ifla65/papers/080-155e.htm> (26-02-2019).
- CLAVEL-MERRIN, G. (2004). MACS (Multilingual Access to Subjects): a virtual authority file across languages. *Cataloging & Classification Quarterly*, 39 (1-2), 323-330, DOI: 10.1300/J104v39n01\_02
- COHEN, A. (2017). "Each of us has a name": creating a multilingual, multiscript national authority database. *Catalogue & Index*, 186, 15-19 [https://archive.cil-ip.org.uk/sites/default/files/media/document/2017/ci186cohen\\_ahava.\\_each\\_of\\_us\\_has\\_name\\_creating\\_a\\_multilingual\\_multis.pdf](https://archive.cil-ip.org.uk/sites/default/files/media/document/2017/ci186cohen_ahava._each_of_us_has_name_creating_a_multilingual_multis.pdf) (26-02-2019).
- GROSJEAN, F. (1982). *Life with two languages: an introduction to bilingualism*. Cambridge, MA: Harvard University Press.
- LANDRY, P. (2000). The MACS Project: Multilingual Access to Subjects (LCSH, RAMEAU, SWD). In *66th IFLA Council and General Conference, Jerusalem, August 13-18, 2000*, <http://archive.ifla.org/IV/ifla66/papers/165-181e.pdf> (26-02-2019).
- LEHTINEN, R. and CLAVEL-MERRIN, G. (1998). Multilingual and multi-character set data in library systems and networks: Experiences and perspectives from Switzerland and Finland. In J.D. Byrum, Jr. and O. Madison (eds.) *Multi-script, multilingual, multi-character issues for the online environment: Proceedings of a Workshop sponsored by the IFLA Section of Cataloguing, Istanbul, Turkey, August 24, 1995*. München: K.G. Saur.
- Library and Archives Canada. (2003). *Bilingual Cataloguing Policy*. <https://www.bac-lac.gc.ca/eng/services/cataloguing-metadata/Pages/bilingual-cataloguing-policy.aspx> (26-02-2019).
- MARC 21 Format for Authority Data. 1999 edition, <https://www.loc.gov/marc/authority/> (26-02-2019).
- MAYR, P. and PETRAS, V. (2008). Cross-concordances: terminology mapping and its effectiveness for information retrieval. In *74th IFLA General Conference and Council, 10-14 August 2008, Québec, Canada*, [http://archive.ifla.org/IV/ifla74/papers/129-Mayr\\_Petras-en.pdf](http://archive.ifla.org/IV/ifla74/papers/129-Mayr_Petras-en.pdf) (26-02-2019).
- Multilingual Authority Records in the MARC 21 Authority Format*. (2001). <https://>

- [www.loc.gov/marc/marbi/2001/2001-dp05.html](http://www.loc.gov/marc/marbi/2001/2001-dp05.html) (26-02-2019).
- RDA, *Resource Description and Access*, <http://access.rdatoolkit.org/> (26-02-2019).
- RIGBY, C. (2008). *Improving access to aboriginal language materials in the Unicode age: A Nunavut case study in developing multilingual and multiscript descriptive cataloguing standards for an integrated library system*. A dissertation submitted in partial fulfilment of the requirements of the Master of Science Degree in Information and Library Studies. The Robert Gordon University. <https://nunavutlibraryassociation.files.wordpress.com/2012/09/crigbydissfinal.pdf> (26-02-2019).
- Virtual International Authority File (VIAF), <http://viaf.org/> (26-02-2019).
- WERLING, P. (2015). New Dewey Software Features. In *EDUG Symposium, April 17, 2015*, <http://edug.pansoft.de/tiki-index.php?page=2015+meeting> (26-02-2019).

# VIŠEJEZIČNOST U SUSTAVIMA ZA PRONALAŽENJE INFORMACIJA SLJEDEĆI IZAZOV

## KLJUČNE RIJEČI:

*višejezični katalozi, dvojezični katalozi, jezik katalogiziranja, višejezične datoteke autoriziranih podataka imena*

## SAŽETAK

Višejezični sustav za pronalaženje informacija je onaj što kao jedinstveni sustav služi korisničkoj populaciji koja kao cjelina funkcionira na više jezika. Višejezični dizajn sustava za pronalaženje informacija nadoknađuje razlike u jezičnim preferencijama krajnjih korisnika. U idealnom slučaju, sustav bi trebao biti dobar na svim jezicima koje nudi. Mnoge razine u katalogu imaju jezičnu orijentaciju i potrebno ih je uzeti u obzir pri preusmjeravanju jednojezičnog kataloga u višejezični katalog. Tu se ubrajaju: korisničko sučelje, konvencije prikaza podataka i svi metapodaci, od podataka preuzetih s jedinice građe do podataka koje dodaje knjižničar, pristupnica i predmetnih jezika. Unatoč poznatim problemima, u širokoj su uporabi sučelja koja se mogu prebacivati na više jezika. Metapodaci nisu uvijek jezično neutralni. Izbor jezika podrazumijeva se u postojećem kataložnom pravilniku ili smjernicama. Donošenje kriterija na temelju kojih se utvrđuju uvriježeni oblici imena ili naslova na prioritetnom jeziku kataložne agencije, rezultira opisima koji su usmjereni na korisnike određenog jezika. Prioritetnim oblicima upravljaju datoteke autoriziranih podataka imena, koje se mogu upotrijebiti za prilagođavanje bibliografskih podataka prikazanih krajnjem korisniku. Metode za predmetni pristup odnose se na izravno povezivanje različitih predmetnih shema ili primjenu klasifikacije kao zajedničkog čvorišta za navigaciju među jezicima. Unatoč napretku koji je postignut, i dalje postoje osnovni izazovi pružanja potpuno funkcionalnih dvojezičnih i višejezičnih sustava. Unatoč tomu, možda je ipak korisnije imati djelomičnu višejezičnost nego da se ona uopće ne pokuša uvesti.